

1 **Rapid classification and prediction of COVID-19 severity by MALDI-TOF**
2 **mass spectrometry analysis of serum peptidome**

3

4 **Running title:** MALDI-TOF MS to classify and predict COVID-19 severity

5

6 **Authors:** Rosa M. Gomila,^a Gabriel Martorell,^a Pablo A. Fraile-Ribot,^{b,c} Antonio
7 Doménech-Sánchez,^{c,d,4} Antonio Oliver,^{b,c} Mercedes García-Gasalla,^{c,e}
8 Sebastián Albertí^{a,c,d,#}

9

10 **Author's affiliation:**

11 ^aServicio Científico-Técnico, Universidad de las Islas Baleares, Palma de
12 Mallorca, Spain.

13 ^bServicio de Microbiología. Hospital Universitario Son Espases, Palma de
14 Mallorca, Spain.

15 ^cInstituto de Investigación Sanitaria de les Illes Balears, Palma de Mallorca,
16 Spain.

17 ^dInstituto Universitario de Investigación en Ciencias de la Salud, Universidad de
18 las Islas Baleares, Palma de Mallorca, Spain.

19 ^eServicio de Medicina Interna. Hospital Universitario Son Espases, Palma de
20 Mallorca, Spain.

21

22 # Address correspondence to Sebastian Alberti, sebastian.alberti@uib.es.

23

24 **ABSTRACT**

25 Classification and early detection of severe COVID-19 patients is urgently
26 required to establish an effective treatment. Here, we tested the utility of matrix-
27 assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-
28 TOF MS) to classify and predict the severity of COVID-19 in a clinical setting.
29 We used this technology to analyse the mass spectra profiles of the sera from
30 80 COVID-19 patients, clinically classified as mild (33), severe (26) and critical
31 (21), and 20 healthy controls. We found a clear variability of the serum
32 peptidome profile depending on COVID-19 severity. Seventy-eight peaks were
33 significantly different and 12 at least four fold more intense in the set of critical
34 patients than in the mild ones. Analysis of the resulting matrix of peak
35 intensities by machine learning approaches classified severe (severe and
36 critical) and non-severe (mild) patients with a 90% of accuracy. Furthermore,
37 machine learning predicted correctly the favourable outcome of the severe
38 patients in 85% of the cases and the unfavourable in 38% of the cases. Finally,
39 liquid chromatography mass spectrometry analysis of sera identified five
40 proteins that were significantly upregulated in the critical patients. They included
41 serum amyloid proteins A1 and A2, which probably yielded the most intense
42 peaks with m/z 11,530 and 11,686 detected by MALDI-TOF MS.
43 In summary, we demonstrated the potential of the MALDI-TOF MS as a bench
44 to bedside technology to aid clinicians in their decisions to classify COVID-19
45 patients and predict their evolution.

46 INTRODUCTION

47 Coronavirus infectious disease 19 (COVID-19) was first reported in Wuhan,
48 Hubei province, China as a new coronavirus disease caused by a positive-
49 strand RNA virus designated as severe acute respiratory syndrome coronavirus
50 2 (SARS-CoV-2) (1). This virus is responsible of a pandemic of unprecedented
51 dimensions. Approximately 80% of COVID-19 cases are asymptomatic or
52 present mild symptoms, such as fever, cough, fatigue, and dyspnea.
53 Conversely, about 20% of patients with COVID-19 develop viral pneumonia,
54 with an exaggerated host inflammatory response and hypoxia, requiring
55 intubation and mechanical ventilation (2,3). These patients, classified as
56 clinically severe or critical life threatening infections, are mainly diagnosed
57 empirically based on a set of clinical characteristics. However, patients with
58 these symptoms have already evolved to a serious clinical condition that
59 requires specialized intensive care. Therefore it is essential to set up novel and
60 rapid approaches to identify biomarkers for symptom onset and disease
61 progression to facilitate triage of patients and establish appropriate treatments.
62 Peptidome-based studies using serum from patients and high-throughput
63 spectrometric techniques promise to be valuable for the identification of COVID-
64 19-associated biomarkers. Serum may contain proteins induced by the
65 systemic effects or released to the lung as a result of the viral infection. Thus,
66 patient serum can reflect the physiological or pathological state. Indeed, a
67 proteomic and metabolomic analysis of serum from 46 COVID-19 patients
68 performed by Shen et al demonstrated that using serum proteins and
69 metabolite biomarkers it is possible, not only classify patients according to their
70 grade of severity, but also predict the progression to severe COVID-19 (4).

71 More recently, Messner et al re-designed a high-throughput mass spectrometry
72 platform that enabled the identification of up to 27 potential biomarkers that
73 were differentially expressed depending on the severity grade of COVID-19 (5).
74 Although the technologies used in both studies are highly sensitive and provide
75 robust results, they are time consuming, requires specialized personnel and
76 most importantly, they are not available in most of the hospitals, so their
77 translation bench to bedside is limited.

78 In the present study, we used MALDI-TOF MS, a simple and fast technology,
79 available in most of the hospitals, to conduct a comparative analysis of serum
80 from 80 COVID-19 patients. Our results demonstrate the value and power of
81 MALDI-TOF to classify and predict the progression of COVID-19 in a clinical
82 setting.

83

84 **MATERIALS AND METHODS**

85 **Ethics Statement**

86 The study was conducted in accordance with the Declaration of Helsinki. All
87 human samples were taken after written consent of the participants. They were
88 informed of the purposes of the study, which was approved by the Ethics
89 Review Board of the Illes Balears (CEI).

90

91 **Participants**

92 This study included a total of 80 COVID-19 patients who attended Hospital Son
93 Espases, the reference hospital of the Balearic Islands, between March 2020
94 and June 2020. COVID-19 cases were confirmed based on the Chinese
95 management guideline for COVID-19 (6). Only patients with a positive RT-PCR
96 test were enrolled.

97 The severity grade of COVID-19 was defined based on the abovementioned
98 guideline (6). Accordingly, COVID-19 patients were classified into three
99 subgroups: mild, severe, or critical. Mild included non-pneumonia and mild
100 pneumonia cases. Severe was characterized by dyspnea, respiratory frequency
101 ≥ 30 /minute, blood oxygen saturation $\leq 93\%$, $\text{PaO}_2/\text{FiO}_2$ ratio < 300 , and/or lung
102 infiltrates $> 50\%$ within 24–48 hours. Critical cases were those that exhibited
103 respiratory failure, septic shock, and/or multiple organ dysfunction/failure.

104 Twenty healthy volunteers, including 13 recovered from COVID-19 were also
105 included in the study.

106 **Sample collection**

107 Blood samples were collected into anticoagulant free-tubes. The tubes were
108 centrifuged at 2,500 rpm at 20°C for 10 min within a 30-min time frame. Serum
109 from each patient sample was then collected, aliquoted and stored at –80°C.
110 Each serum sample was heat inactivated a 56°C for 90 min prior to analysis.

111

112 **MALDI-TOF serum sample preparation**

113 The preparation of the serum and the analysis of the samples by MALDI-TOF
114 MS were performed as previously described (7). Serum samples were purified
115 and concentrated using reversed phase C18 tip, Pierce™ C18, following the
116 manufacture instructions. In brief, 20 µl of serum were mixed with 15 µl of
117 trifluoroacetic acid (TFA) 2%. Samples were passed through the Pierce™ C18
118 tip by pipetting up and down repeatedly (20 times), followed by the separation
119 of the unbound protein solution. After washing the Pierce™ C18 tip with 10 µl of
120 0.1% TFA, the bound proteins/peptides were eluted with 4 µl of 0.1%
121 TFA:CH₃CN (1:1, v/v). The solution was passed through the Pierce™ C18 tip
122 repetitively (8-10 times). The eluted protein/peptide solution was mixed with 4 µl
123 of a-cyano-4-hydroxycinnamic acid (CHCA) matrix solution (10 mg of CHCA in
124 1 ml of 5% TFA:CH₃CN, 1:1, v/v) and 1.5 µl of this mixture were spotted onto a
125 MTP 384 target plate ground steel (Bruker Daltonics, Leipzig, Germany) and
126 overlaid with 2 µl of CHCA matrix and allowed to air-drying. The analysis of
127 each sample was conducted in triplicate.

128

129 **MALDI-TOF analysis**

130 Measurements were performed on an Autoflex III MALDI-TOF mass
131 spectrometer (Bruker Daltonics, Leipzig, Germany) equipped with a 200 Hz

132 smart beam laser. Spectra were generated by averaging 1000 single laser
133 shots (100 shots at 10 different spot positions) at a laser frequency of 200 Hz
134 and detected in linear positive mode. The IS1 voltage was 20.1 kV, the IS2
135 voltage was maintained at 18.7 kV, the lens voltage was 8.4 kV, and the
136 extraction delay time was 140 ns. Peaks between 2,000 and 25,400 Da were
137 selected for analysis. Mass accuracy was calibrated externally using the Bruker
138 Bacterial Test Standard. Triplicates of each sample were obtained.

139

140 **MALDI-TOF mass data processing**

141 Raw mass spectra obtained by MALDI-TOF MS was analysed using the
142 MALDIquant R package (8). Square root transformation, peak smoothing,
143 baseline correction, and intensity normalization were performed on each mass
144 spectrum. The average spectrum from the triplicates was obtained. Peaks were
145 detected and binned across all average spectra with a signal to noise ratio of 5
146 and a tolerance of 0.002. Peaks presents in less than 25% of the spectra were
147 rejected. All spectra from the groups under study were pre-processed, and peak
148 detection was applied to obtain an intensity matrix. The resulting matrix of peak
149 intensities was used for Principal Component Analysis (PCA) (9) and Machine
150 learning (ML) approaches (10). To set up ML, the radial basis function was used
151 as the kernel function and ANOVA was used to select the 45 most relevant
152 peaks.

153

154 **Liquid chromatography-mass sample preparation**

155 Three microliters of serum were diluted to 1 ml with 50 mM Ammonium
156 bicarbonate (0.2 $\mu\text{g}/\mu\text{l}$ taking the average of plasma proteins as 80 mg/ml). One

157 hundred μL of the dilution were reduced with 11 μl of 50 mM dithiothreitol for 30
158 minutes at 56°C and were alkylated with 12.5 μl of indole-3-acetic acid 20 mM
159 for 20 minutes in the dark at 37 °C. Total volume 123.5 μl , containing likely 24
160 μg of plasma protein, were digested with 10 μL of trypsin 100 ng/ μl at 37°C
161 overnight. Ten microliters of formic acid (FA) 5% were added to stop the
162 digestion.

163

164 **Liquid chromatography-mass setup**

165 The digested peptides were analysed by liquid chromatography mass
166 spectrometry (LC-MS/MS) with a nanoflow Agilent series 1200 LC system
167 (Agilent Technologies, Waldbronn, Germany), with an autosampler equipped
168 with an 8 μl capillary loop, coupled to a Q-Exactive Plus Hybrid Quadrupole-
169 Orbitrap mass spectrometer (ThermoFisher®Scientific) in data dependent
170 acquisition mode. For each acquisition, peptides were loaded onto a precolumn
171 (ZORBAX 300 SB-C18, 5 μm , 5 mm *0.3 mm i.d.) at a flow rate of 15 $\mu\text{l}/\text{min}$ for
172 2 min and then analysed using a 235 min LC gradient (from 3% to 97% buffer
173 B) at a flow rate of 250 nl/min (analytical column, ZORBAX 300 SB-C18, 3.5
174 μm , 150 mm *0.075 mm i.d.). Buffer A was H₂O containing FA 0.1%, and buffer
175 B was acetonitrile with 0.1% FA. All reagents were MS grade. The *m/z* range of
176 MS1 was 350-1,650 with the resolution at 140,000 (at 200 *m/z*), automatic gain
177 control (AGC) target of $3e^6$, and maximum ion injection time (max IT) of 250 ms.
178 Top 10 precursors were selected for MS/MS experiment, with a resolution at
179 17,500 (at 200 *m/z*), AGC target of $5e4$, and max IT of 200 ms. The isolation
180 window of selected precursor was 4 *m/z*.

181

182 **Liquid chromatography-mass data processing**

183 The resultant mass spectrometric data were analysed using Proteome
184 Discoverer (Version 2.2.0.388, Thermo Fisher Scientific) using a protein
185 database composed of the Homo sapiens fasta database downloaded from
186 UniProtKB on 12 Jul 2020, containing 20,304 reviewed protein sequences, and
187 the SARS-CoV-2 virus fasta downloaded from UniProtKB on 20 May 2020,
188 containing 13 protein sequences. Enzyme was set to trypsin with four missed
189 cleavage tolerance. Static modifications were set to carbamidomethylation
190 (+57.02146) of cysteine and variable modifications were set to oxidation
191 (+15.99492) of methionine and dimethylation (+28.03075) of peptides N-termini.
192 Precursor ion mass tolerance was set to 10 ppm, and product ion mass
193 tolerance was set to 0.6 Da. The peptide-spectrum-match allowed 1% target
194 false discovery rate (FDR) (strict) and 5% target FDR (relaxed). Normalization
195 was performed against the total peptide amount. The other parameters followed
196 the default setup.

197

198 RESULTS

199 We acquired MALDI mass spectra of 82 serum samples obtained from 80
200 different COVID-19 patients (from two patients, we analysed two samples
201 (109/144 and 143/141) collected at different severity grades); 34 samples were
202 collected from mild COVID-19 patients, 26 from severe patients and 22 from
203 critical patients. We also analysed 20 serum samples obtained from healthy
204 people, including 13 samples collected from individuals recovered from COVID-
205 19. The relevant characteristics of each group are shown in Table 1.

206 All spectra from the four groups under study were processed and peak
207 detection was applied to obtain an intensity matrix of 135 peaks in the mass
208 range of 2,000 to 25,000 daltons using MALDIquant. To select the most
209 characteristic peaks distinguishing the groups classified according to COVID-19
210 severity, we applied Post hoc Turkey's HSD analysis, which identified 78 peaks
211 with a FDR (False Discovery Rate) value ≤ 0.05 . Figure 1A illustrates the
212 quantitative variability of those peaks, significantly different, that exhibited a log
213 fold change ≥ 2 for COVID-19 severity on a heatmap. We found clear
214 differences between critical and mild patients. Twelve peaks were significantly
215 different and at least four fold more intense in the set of critical patients than in
216 the mild ones (Figure 1B). However, only two peaks (m/z 11,530 and 11,686)
217 were significantly different between severe and critical patients (Figure 1B).

218 These results encouraged us to apply ML approaches to classify and predict
219 COVID-19 severity. We built two support vector machine learning models using
220 the 135 peaks (m/z) obtained from each sample by MALDI-TOF MS. In the first
221 model, patients were classified as severe and non-severe. Severe patients
222 group included those that required oxygen support (severe and critical patients),

223 while non-severe were those without oxygen support (mild patients). Seventy-
224 five percent of the samples from each group were randomly selected to form
225 the training set, while the remaining 25% of the samples were used for
226 validation as an independent test cohort. This process was repeated five
227 independent times, every time ML was tested through fivefold cross-validation.
228 This model reached an average area under curve (AUC) of 0.911 for Receiver
229 Operating Characteristic (ROC) curve (8).

230 Using this model, 90 % of the samples were correctly classified with only a 5%
231 of false positives and 5% of false negatives. PCA separated individuals
232 according to the severity of COVID-19 (Figure 2) and helped to identify some
233 outliers (samples 4, 11, 30, 34 from mild patients) previously detected in the
234 heatmap and in the bloxplots.

235 Next, we assessed whether ML would be able to predict clinical evolution of
236 severe patients (grade 2). For this model, a database constructed using the
237 mass spectra of samples from mild (grade 1) and critical (grade 3) patients,
238 excluding outliers samples, was challenged with the mass spectra of samples
239 from severe patients (grade 2). As above, this process was repeated five
240 independent times, every time ML was tested through fivefold cross-validation.
241 This model reached an AUC of 0.956 for ROC curve. Green and red brackets
242 below heatmap panel in figure 1A indicate those clinically severe (grade 2)
243 patients that were clustered by ML as mild or critical, respectively. All severe
244 patients that were classified by ML as mild had a good clinical evolution except
245 patients of samples 108 and 144, which worsened to critical 8 days later,
246 suggesting that this technology is quite accurate to predict favourable
247 prognostics. On the other hand, five patients (samples 111, 143, 101, 118 and

248 137) classified as clinically severe (grade 2) when the samples were collected,
249 which 48-96 h later evolved to critical (grade 3), were clearly clustered in the
250 group of critical patients using machine learning, while the remaining 8 patients
251 (70, 73, 66, 67, 62, 59, 53, 58) did not worsen to grade 3.

252 Most of the discriminating peaks identified by MALDI-TOF MS analysis had a
253 low molecular weight (< 5,000 Da). They probably resulted from the
254 fragmentation of proteins upregulated in severe and critical patients.
255 Interestingly, the most substantial intensity difference were exhibited by the
256 peaks with m/z of 11,530 and 11,686, which might correspond to unfragmented
257 proteins of the acute phase induced by the virus. To investigate this hypothesis,
258 we performed a proteomic analysis of the samples by LC MS/MS. We identified
259 five proteins that were significantly upregulated according to the severity of the
260 disease; the serum amyloid A2 protein (SAA2), the C reactive protein (CRP),
261 the serum amyloid protein A1 (SAA1), the lipopolysaccharide binding protein
262 (LBP) and the gamma chain of the fibrinogen (FGG). Figure 3A illustrates the
263 quantitative variability of these proteins on a heatmap. Only the serum level of
264 SAA2 exhibited significant increments between mild and severe patients and
265 between severe and critical patients (Figure 3B). In addition, SAA1, CRP, LBP
266 and FGG were increased in the serum from the critical patients compared with
267 to mild patients, while only CRP and SAA1 were increased in the severe
268 patients compare to mild patients (Figure 3B). Given that the molecular weight
269 of SAA1 and SAA2 is approximately 11.7 kDa, depending on the isoform (11),
270 and that we found a good correlation between the level of both proteins and the
271 intensity of the peaks with m/z of 11,530 and 11,686, we suggest that these
272 peaks might correspond to the serum amyloid proteins A1 and A2.

273 **DISCUSSION**

274 In this study we demonstrate that the molecular changes that occur in the sera
275 of COVID-19 patients may be detected by MALDI-TOF MS analysis generating
276 peptidome profiles that may be used as clinical classifiers. In addition, we show
277 that it is possible to predict the progression of the disease using the peptidome
278 signatures obtained with this technology. Finally, we provide strong evidences
279 that serum amyloid A2 protein is one of the major biomarkers of severe COVID-
280 19 disease.

281 To our knowledge, only two previous studies reported the use of mass
282 spectrometry analysis of serum from COVID-19 patients to classify disease
283 severity (4, 5). However, both studies were performed using sophisticated
284 technologies, which are not available in most of the hospitals. Our challenge
285 was to test whether MALDI-TOF MS analysis, a simpler technology available in
286 most of the clinical microbiology laboratories for identification of microbial
287 species, was able to achieve similar results. Our study classified severe
288 patients with a high accuracy (90%), very similar to that obtained in a previous
289 report (93%) (4), with a very low number of false results.

290 We represented the changes of those peaks that exhibited major changes upon
291 grouping the patients according to CDC Chinese severity criteria, ranging from
292 scale 1 to scale 3 in a heatmap, which graphically illustrated how level changes
293 in these peaks reflected a progression from mild to critical COVID-19.
294 Interestingly, our peptidome profile data identify the most important changes
295 within the severe patients, upon which a patient is put on oxygen supply. This
296 observation is consistent with the proteome analysis conducted by Messner et
297 al, who found that at molecular level the requirement of oxygen supply

298 coincided with the progression to severe disease (5). In contrast, mild patients
299 have a peptidome signature virtually identical to the healthy controls suggesting
300 that in non-severe patients changes are restricted to the site of infection, the
301 respiratory tract, without significant molecular systemic alterations. Bloxplot
302 representation also identified outlier samples that were confirmed by PCA. As
303 occurred in the study by Messner et al. (5), it is likely that these samples were
304 collected from patients with certain underlying pathologies or under specific
305 treatments that altered the peptidome signature.

306 Case studies demonstrated the clinical utility of the peptidome profiles to
307 classify and predict COVID-19 evolution. First, five samples from patients
308 classified as clinically severe, presented a profile very similar to the group of
309 critical patients in the heatmap and clustered in this group using ML 48-96 h
310 before they clinically progressed to critical. Second, sample 141 obtained from
311 the same patient as sample 143, which evolved from severe to critical, was
312 collected two days before it was discharged and presented a peptidome profile
313 clustered as mild by ML and in the heatmap. On the other hand, samples from
314 patients completely recovered of COVID-19 (blue bracket in figure 1A) or those
315 that evolved from severe to mild had peptidome signatures similar to those
316 observed in the healthy control patients. Conversely, peptidome profile of
317 sample 144, obtained from the same patient as sample 109, but eight days
318 before he progressed to critical, was not able to predict the clinical trajectory of
319 this patient.

320 Overall, these results suggest that peptidome profiles obtained by MALDI-TOF
321 MS may represent a good prognostic tool to support clinical decisions.
322 However, it would be interesting to conduct a longitudinal study with sequential

323 daily samples from a cohort of patients at different grades of severity until their
324 recovery to assess the anticipation time of prediction.

325 One of the potential limitations of our study is that due to the rapid response
326 required in the initial stages of the pandemic situation, we collected samples
327 from the patients that were admitted in our hospital using as unique criteria that
328 they were hospitalized due to a SARS-CoV-2 infection. Therefore our study did
329 not take in account some confounding factors, like age. Nonetheless, the
330 change of the intensity of the peaks between groups substantially exceeded the
331 variability observed within each group with ages ranging from 33 to 89,
332 suggesting that differences in the peptidomes profiles of different groups are
333 poorly influenced by confounding factors.

334 Two of the most intense peaks detected in the sera from critical patients had
335 *m/z* of 11,530 and 11,686 that might correspond to two different isoforms of the
336 serum amyloid A protein (11). This acute phase markers, induced by the
337 proinflammatory cytokine IL-6, were two of the predominant proteins detected
338 by both Shen et al and Messner et al in their respective studies (4,5). As in our
339 study, they also detected CRP, LBP and FGG as clear protein biomarkers for
340 COVID-19 severity.

341 In conclusion, our study supports the potential of the MALDI-TOF MS as a fast
342 and clinically available technology to aid clinicians in their decisions on COVID-
343 19 patients and identifies serum amyloid protein A2 as an excellent biomarker
344 to monitor COVID-19 patients.

345

346 **ACKNOWLEDGMENTS**

347 The authors thank Sara Fernández for helpful discussions on statistical

348 analysis. This study was funded by Instituto de Investigación Sanitaria de las

349 Islas Baleares (IdISBa).

350

351 **REFERENCES**

352

353 1. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Zhang L, Fan G, Xu J, Gu X,
354 Cheng Z, Yu T, Xia J, Wei Y, Wu W, Xie X, Yin W, Li H, Liu M, Xiao Y, Gao
355 H, Guo L, Xie J, Wang G, Jiang R, Gao Z, Jin Q, Wang J, Cao B. 2020.
356 Clinical features of patients infected with 2019 novel coronavirus in Wuhan,
357 China. *Lancet* 395: 497-506.

358

359 2. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, Qiu Y, Wang J, Liu Y, Wei
360 Y, Xia J, Yu T, Zhang X, Zhang L. 2020. Epidemiological and clinical
361 characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan,
362 China: a descriptive study. *Lancet* 395: 507-13.

363

364 3. Wang D, Hu B, Hu C, Zhu F, Liu X, Zhang J, Wang B, Xiang H, Cheng Z,
365 Xiong Y, Zhao Y, Li Y, Wang X, Peng Z. 2020. Clinical characteristics of
366 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia
367 in Wuhan, China. *JAMA* 323: 1061-69.

368

369 4. Shen B, Yi X, Sun Y, Bi X, Du J, Zhang C, Quan S, Zhang F, Sun R, Qian
370 L, Ge W, Liu W, Liang S, Chen H, Zhang Y, Li J, Xu J, He Z, Chen B, Wang
371 J, Yan H, Zheng Y, Wang D, Zhu J, Kong Z, Kang Z, Liang X, Ding X, Ruan
372 G, Xiang N, Cai X, Gao H, Li L, Li S, Xiao Q, Lu T, Zhu Y, Liu H, Chen H,
373 Guo T. 2020. Proteomic and metabolomic characterization of COVID-19
374 patient sera. *Cell* 182:59-72.

375

- 376 5. Messner CB, Demichev V, Wendisch D, Michalick L, White M, Freiwald A,
377 Textoris-Taube K, Vernardis SI, Egger AS, Kreidl M, Ludwig D, Kilian C,
378 Agostini F, Zelezniak A, Thibeault C, Pfeiffer M, Hippenstiel S, Hocke A,
379 von Kalle C, Campbell A, Hayward C, Porteous DJ, Marioni RE,
380 Langenberg C, Lilley KS, Kuebler WM, Mülleder M, Drosten C, Suttorp N,
381 Witzenrath M, Kurth F, Sander LE, Ralser M. 2020. Ultra-high-throughput
382 clinical proteomics reveals classifiers of COVID-19 infection. *Cell Syst*
383 11:11-24.
- 384
- 385 6. The novel coronavirus pneumonia emergency response epidemiology
386 team. The epidemiological characteristics of an outbreak of 2019 novel
387 Coronavirus diseases (COVID-19) — China, 2020. *China CDC Weekly*
388 2020; 2: 113-122.
- 389
- 390 7. Barceló F, Gomila R, de Paul I, Gili X, Segura J, Pérez-Montaña A,
391 Jimenez-Marco T, Sampol A, Portugal J. 2018. MALDI-TOF analysis of
392 blood serum proteome can predict the presence of monoclonal
393 gammopathy of undetermined significance. *PLoS One* 13:e0201793.
- 394
- 395 8. RStudio Team. *RStudio: Integrated Development Environment for R*.
396 Boston, MA. Retrieved from <http://www.rstudio.com/> 2015.
- 397
- 398 9. Wold S, Esbensen K, Geladi, P. Principal component analysis.
399 *Chemom.Intell. Lab. Syst* 1987; 2:37–52.
- 400

- 401 10. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, Mann M,
402 Cox J. 2016. The Perseus computational platform for comprehensive
403 analysis of (prote)omics data. *Nat Methods* 13:731-40.
404
- 405 11. Li J, Xie Z, Shi L, Zhao Z, Hou J, Chen X, Cui Z, Xue P, Cai T, Wu P, Guo
406 S, Yang F. 2012. Purification, identification and profiling of serum amyloid A
407 proteins from sera of advanced-stage cancer patients. *J Chromatogr B
408 Analyt Technol Biomed Life Sci* 15;889-890.
409
410

411 **Table 1. Relevant characteristics of the individuals included in this study**
412

	Healthy (n=20)	COVID-19 patients		
		Mild (n=33)	Severe (n=26)	Critical (n=21)
Sex				
Male	11	11	16	21
Female	9	22	10	0
Age				
Mean ± SD	50±12.3	55.2±16.7	63.2±9.9	63.2±10.1
Median	50.5	49	65	65.5
Range	27-73	34-89	47-81	33-76

413

414 **FIGURE LEGENDS**

415 **Figure 1. MALDI mass spectra of serum indicate clinical severity in**

416 **COVID-19.**

417 (A) Heatmap illustrates peptidome profiles that inform on COVID-19 severity.

418 Heatmap was generated using the ComplexHeatmap package (8). Groups were

419 classified according to COVID-19 severity following the Chinese management

420 guideline for COVID-19. Blue bracket below heatmap indicates healthy

421 individuals recovered of COVID-19, while green and red brackets indicate

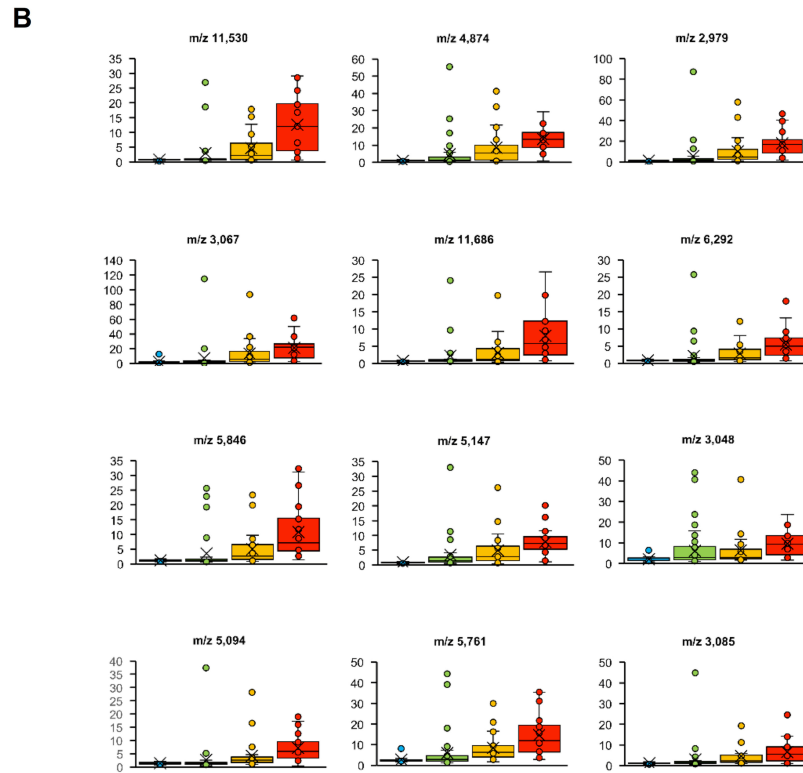
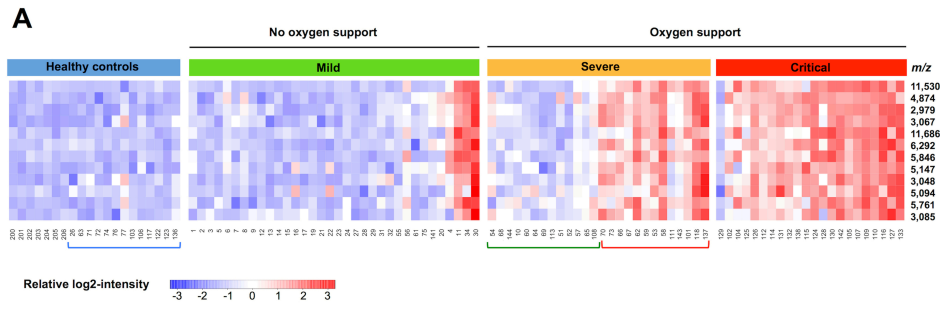
422 samples from patients classified by ML as mild or critical, respectively.

423 (B) Peaks with increased intensity depending on COVID-19 severity. The boxes

424 show the first and third quartiles as well as the median (middle), the mean

425 (cross), and the outliers (circles outside the whiskers).

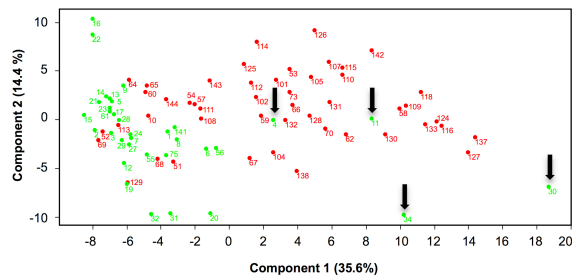
426



427
428

429 **Figure 2.** PCA of the mass spectra of the samples from the non-severe (mild)
430 (green dots) and severe (severe and critical) (red dots) patients. Arrows indicate
431 outlier samples.

432

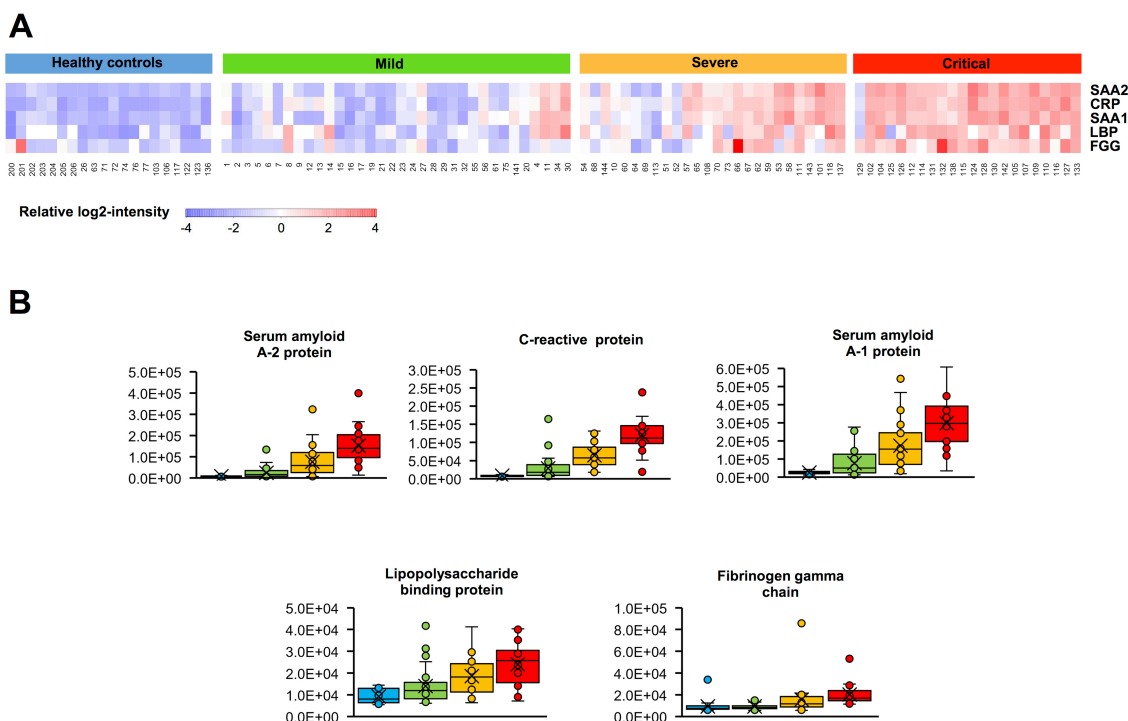


433

434

435 **Figure 3. Upregulated proteins according to COVID-19 severity.**

436 (A) Heatmap illustrates proteins that inform on COVID-19 severity. Heatmap
437 was generated using the ComplexHeatmap package (8). Only proteins that
438 were present in more than 70% of the samples identified by at least 5 peptides
439 and significantly increased ($P < 0.05$) by a log-2 fold change > 2 were included in
440 the figure. Groups were classified according to COVID-19 severity following the
441 Chinese management guideline for COVID-19. SAA2 (serum amyloid A2
442 protein), CRP (C reactive protein), SAA1 (serum amyloid protein A1), LBP
443 (lipopolysaccharide binding protein) and FG (gamma chain of the fibrinogen).
444 (B) Upregulated proteins depending on COVID-19 severity. The boxes show the
445 first and third quartiles as well as the median (middle), the mean (cross), and
446 the outliers (circles outside the whiskers).
447



448